

Re: Speeding up an application – general rules

Source: <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2006-12/msg01419.html>

- *From:* Eric Schwartz <emschwar@xxxxxxxxxx>
 - *Date:* 21 Dec 2006 22:15:02 -0700
-

"Petyr David" <phynkel@xxxxxxxxxx> writes:

Basically: the script uses perl's system command to run a long winded "find" command which is piped to sed to correct patterns that match HTML markers.

You are unclear here, which is why we generally ask you to post example code. In fact, it's really kinda hard to say anything for sure because you didn't. I'm not sure, for instance, if you pipe the output of find to sed, or if you iterate over the list of files returned by find and run sed on the contents of those files. I'm guessing the former, but it's just a guess. If you want people to be able to help you the best way possible, you probably don't want to make them guess.

The matching lines are then shoved into an array.

Which lines? Are you talking about contents of the files, or names of files? Now I think you're talking about contents. It would help if you were more clear.

The elements of the array are moved into a hash for the purpose of sorting the file names.

Er, now I think you're talking about file names.

Then file names and matching lines are printed.

Now I have no idea. What are you actually doing? Can you please show some code?

Re: Speeding up an application – general rules

Q: Can I speed things by eliminating the sed command and letting Perl filter and modify the matching patterns? If so, how much of a performance gain?

Honestly, rather than asking us, you should ask Perl. The answer to "how do I speed things up?" is profile profile profile! Until you profile, you don't know what will help.

'perldoc -q profile' mentions the Devel::DProf module, and you can use 'perldoc Devel::DProf' to find out more about it. You'll also want to learn about the Benchmark module ('perldoc Benchmark'), which will help you compare two different ways of doing the same thing to find out which is faster.

Is using Perl's grep to search through every file for the pattern faster than using the find command?

Wait, are you on Windows? It's been a very long time, but I vaguely recall that the Windows 'find' command searches in files, whereas the Unix one mostly just looks at file names and metadata.

The find command has the advantage that I can search for files of a certain date rather easily. Again: could that be done more rapidly by Perl's looking at the file's mod time?

Those questions really depend on such a large a number of things, including your system's OS, configuration, load for other tasks, etc, that it's almost impossible for anyone to tell you for certain. Honestly, even if somebody were to give you an answer here, I wouldn't believe them-- they may be telling you what worked for them, but it might not be the same for you. Profile, then optimize the worst-performing part, then profile again, optimize what's left, and repeat. Take care that in optimizing one part you don't make another slower-- but that's all part of the art, really.

Any thoughts or suggestions would be appreciated

Enjoy. But next time, please post some code, so we can actually tell what you're doing. Making people guess and make stuff up is frustrating for us, because we can't tell if we're guessing right, or going completely off the deep end. I hope I was helpful anyway.

--Eric

.